Transfer Function Model of Physiological Mechanisms Underlying Temporal Visual Discomfort Experienced When Viewing Stereoscopic 3D Images

Taewan Kim, Sanghoon Lee, *Senior Member, IEEE*, and Alan Conrad Bovik, *Fellow, IEEE*

***Abstract*— When viewing 3D images, a sense of visual comfort (or lack of) is developed in the brain over time as a function of binocular disparity and other 3D factors. We have developed a unique temporal visual discomfort model (TVDM) that we use to automatically predict the degree of discomfort felt when viewing stereoscopic 3D (S3D) images. This model is based on physiological mechanisms. In particular, TVDM is defined as a second-order system capturing relevant neuronal elements of the visual pathway from the eyes and through the brain. The experimental results demonstrate that the TVDM transfer function model produces predictions that correlate highly with the subjective visual discomfort scores contained in the large public databases. The transfer function analysis also yields insights into the perceptual processes that yield a stable S3D image.**

***Index Terms*— Visual discomfort, stereoscopic 3D (S3D) image, temporal visual discomfort assessment, temporal visual discomfort model (TVDM), neural activity, accommodation and vergence (A/V) conflict.**

I. Introduction

RECENT advances in 3D display and image processing technology have matched an explosive increase in the demand of 3D content by consumers [1]–[4]. Many Hollywood movies are now produced using 3D technologies [5], and the 2012 London Olympic Games were broadcast in 3D to deliver a more vivid sensation of sport [6]. Simultaneously, a number of important issues have arisen regarding the discomfort felt when viewing 3D content. These issues are of particular interest to high-end service purveyors, 3D display makers and the digital cinema industry [3]–[5]. Unlike the viewing of 2D content, ocular adjustments to 3D depth content can induce neurological symptoms such as eye strain, motion sickness and headaches [1], [2], [7], [8]. The major factors underlying the loss of visual discomfort when viewing 3D content have been widely investigated in video engineering [1], [5], [9], ophthalmology [7], [10] and neurology [11].

One of the most significant factors that affects visual discomfort is binocular parallax, which determines, in part, the degree of dissociation between the accommodation and vergence processes [1]–[3], [7]. When a visually fixated object moves toward the viewer, the eyes converge, causing the lens to thicken and the pupils to constrict. If the object moves away from the viewer, the opposite phenomena occur. Conflicts between accommodation and vergence caused by viewing real world 3D (stereoscopic) images rarely occur. However, when viewing 3D content with implied depths on a flat panel monitor at a fixed distance from the viewer, such conflicts can easily happen [3].

Much work has been applied towards modeling and predicting the degree of visual discomfort felt and the quality of experience when viewing stereoscopic 3D (S3D) images, using both quantitative subjective and automatic objective approaches [4], [8], [9]. Subjective experiments have revealed that various factors such as stereographic scene shooting parameters [5] and the magnitudes and distribution of content disparity [7] play important roles in determining the degree of visual discomfort that is felt. For example, measurements of the range and magnitude of content disparity have been used to automatically predict visual discomfort levels [4], [8], [12]. However, a generalized index that captures both global and local perceptual factors affecting visual discomfort has not yet been systematically developed. In particular, most prior approaches have estimated the degree of visual discomfort at a specific moment of decision.[1] However, visual discomfort accumulates as the viewing time increases. A reliable objective visual discomfort model needs to go beyond snapshot-based approaches [4], [8] and static learning-based methods [12]. Rather, a more reliable 3D visual discomfort model should trace those temporal physiological mechanisms that are relevant to the comfortable viewing of stereoscopic presentations over time.

Manuscript received December 15, 2013; revised March 3, 2015; accepted July 19, 2015. Date of publication July 29, 2015; date of current version August 14, 2015. This work was supported by the Ministry of Science, ICT and Future Planning, Korea, in the ICT Research and Development Program 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Sam Hasinoff.

T. Kim and S. Lee (corresponding author) are with the Department of Electrical and Electronic Engineering, Yonsei University, Seoul 120-749, Korea (e-mail: top.kim@sk.com; slee@yonsei.ac.kr).

A. C. Bovik is with the Laboratory for Image and Video Engineering, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712-1084 USA (e-mail: bovik@ece.utexas.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2015.2462026

[1]For example, in subjective 3D experiments, each scoring decision is made after watching an S3D image for about 10 seconds [13].

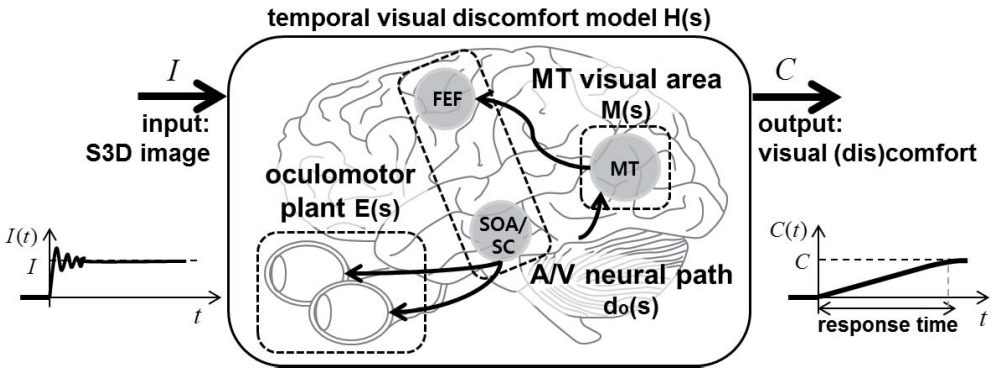


Fig. 1. Visual system model of the temporal discomfort response to an S3D image. The input to the system (I) is sensory depth information from a viewed S3D image and the output of the system (C) is the corresponding degree of experienced visual discomfort. The overall system model is denoted by transfer function $H(s)$ which is composed of the accommodation and vergence (A/V) neural path $d_o(s)$, the oculomotor plant $E(s)$ and the visual area middle temporal (MT) $M(s)$.

A. Motivation of the Temporal Visual Discomfort Model

We describe a new temporal visual discomfort model (TVDM) that is useful for predicting the degree of discomfort felt when viewing S3D images. Of course, there are many factors that contribute to sensations of visual 'discomfort' when viewing S3D images such as window violations, vertical disparities, and intense statistical distribution of texture and color [2]–[4]. Here, we focus only on the conflict that can arise between accommodation and vergence on a flat 3D display, since it is the primary cause of visual discomfort experienced when viewing S3D content [1]–[4]. Inspired by prior models of the behavior of biological vision systems as they respond to dynamic stimuli [14]–[17], we derive and analyze the efficacy of a transfer function (input-output) model that describes the dynamic disparity response and its effect on experienced visual discomfort. The proposed model is based on a theoretical analysis of the physiological response to disparities implied by viewing an S3D image over time.

Other researchers have attempted transfer function analyses of various stages of the visual system [16], [17]. Physiological mechanisms can often be conveniently modeled as a series of processing stages. A transfer function analysis, if accurate, enables useful physical interpretations that can be used to capture the functionality of each stage in the spatio-temporal and spectral domains. Further, the analysis of system behavior by transfer function modeling offers the possibility to simulate the end-to-end system and to make predictions about future responses. Such an approach has been used to model the pupillary light reflex [17] and visual scan paths [18].

Since the flow of visual information is continuous through the eyes and brain, the degree of visual discomfort felt when viewing an S3D image may be regarded as an accumulation of sensory depth information from the viewed 3D scene over time as shown in Fig. 1. Here, we take the simple approach of modeling the relationship between sensory information when viewing an S3D image (I) and the corresponding output degree of experienced visual discomfort (C) as a second-order differential equation, i.e. the overall process is represented as a *second-order system*.

As depicted in Fig. 1, the visual discomfort model consists of the middle temporal (MT) visual area and oculomotor plant, with transfer function $H(s) = C(s)/I(s)$ where $I(s)$ and $C(s)$ are Laplace representations of the input $I(t)$ and the output $C(t)$. The oculomotor plant $E(s)$ dynamically controls eye movements under the control of neural signals delivered by the oculomotor and Edinger-Westphal (EW) neurons. Developing a simple transfer function $M(s)$ describing the time-variant neural responses of visual area MT while viewing an S3D image is a challenging problem. As we will show, however, the '3D visual discomfort' subsystem can be modeled by a second-order transfer function, and this model can be successfully used to predict the temporal sensation of discomfort felt when viewing S3D images.

B. Main Contributions

- We develop a second-order system model that integrates multiple neural structures integral to the perception of S3D images, including the optical nerve, the accommodation and vergence (A/V) neural pathways, the oculomotor plant, and visual area MT. Using a negative feedback loop, the proposed model captures the temporal behavior of visual area MT and the oculomotor plant when viewing an S3D image. A transfer function is developed that can be used to accurately predict the degree of visual discomfort felt when viewing S3D images.
- If the degree of conflict between the accommodation and vergence processes is predicted to increase when viewing an S3D image on a display, then TVDM becomes less stable owing to movement of the system pole. This systematic relationship underlies the predictive power of the system model. Moreover, we show that the perceptual effort required to achieve an acceptable level of visual discomfort increases with the degree of A/V conflict that occurs.
- We analyze the frequency response of visual area MT using Bode plots. In particular, we study the gain and overshoot of the system as A/V conflict is increased.

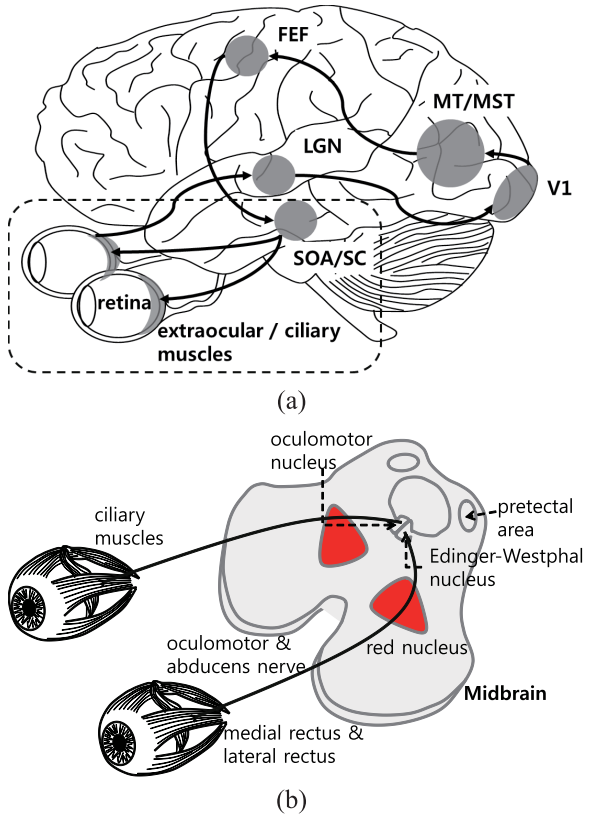


Fig. 2. Brain areas and neural pathways involved in the control of accommodation and vergence responses. (a) Vergence and accommodation is accomplished by control of the extraocular and ciliary muscles, respectively. (b) In the SOA/SC part of midbrain, vergence control of the eyes is affected by the extraocular muscles under control of signaling from the oculomotor nucleus, while the ciliary muscles, which drive accommodation by the crystalline lenses, are controlled by signals emanating from the Edinger-Westphal (EW) nucleus.

II. Temporal Visual Discomfort Model

A. Neural Perception of Accommodation and Vergence

In a natural viewing environment, accommodation and vergence, the two important oculomotor mechanisms, exhibit individual dynamic properties while also maintaining a mutual signal interaction towards achieving a comfortable and realistic viewing experience. Accommodation is the process by which the focussing depth of the crystalline lens of the eye is altered as a function of the distance to a fixated object in order to project a clear and focused image on the retina [16]. Likewise, in order to place the projected image of a fixated object at corresponding retinal points in the two eyes, the process of vergence controls the simultaneous rotation of both eyes in opposite directions [2], [15].

The principal brain areas along the visual pathway that are involved in the control of accommodation and vergent eye movements are shown in Fig. 2 (a). When a scene is projected onto the retina of the eye, it is transduced into parallel streams of neural signals by the photo-receptors. The outputs of spatial neighborhood populations of the retinal photo-receptors are combined via a weighted summation by the post-retinal ganglion cells. The ganglion cells, which have bandpass responses, are thought to affect a form of predictive coding or entropy reduction of the spatial visual signal [20]. The ganglion cell responses then pass via the optic nerve through the lateral geniculate nucleus (LGN), where further processing occurs, including temporal entropy reduction [21]. The LGN responses are then passed *en masse* via the optic radiations to primary visual cortex (area V1), which performs a massively parallel spatio-spectral decomposition of the visual signal into orientation, disparity, motion and frequency tuned bandpass channels [22], [23].

V1 projects the decomposed information onto extrastriate cortical areas including cortical area MT (or V5) and medial superior temporal (MST) cortex. In particular, visual area MT is a central processing stage along the dorsal stream that is deeply implicated in the processing of signals for motion perception and binocular disparity computation [23]. Although many early studies on binocular disparity processing have focused on V1, recent studies indicate that MT plays a major role in subsequent disparity processing and that disparity selectivity there is considerably stronger than in other cortical areas, such as V1 or V4 [24], [25]. In addition, MT neurons directly feed MST neurons, whose collective activity carries substantial information regarding the initiation of vergence eye movements. The responses of MT neurons are deeply involved in the depth perception, as well as being implicated in eliciting vergence eye movements [23], [26], [27].

The frontal eye field (FEF), which is interconnected with signaling from areas MT/MST directs the activation of saccadic, smooth pursuit and vergence eye movements. In addition, the FEF generates motor control signals and carries them to the premotor neurons of the supraoculomotor area (SOA) and to the superior colliculus (SC) of the midbrain where ocular motor neurons are activated for fast and slow vergence, respectively [28], [29].

Finally, the eyes are rotated by the extraocular muscles under the control of signaling from the oculomotor nucleus to achieve vergence, while the ciliary muscles, which force the crystalline lenses to accommodate, are controlled by signals emanating from the EW nucleus, as shown in Fig. 2 (b) [30]. The extraocular muscles, specifically, the medial and lateral recti, are controlled by signals carried by the oculomotor and abducens nerves to achieve convergence and divergence, respectively.

The severity of visual discomfort that is experienced when viewing S3D presentations generally varies over time. The degree of visual discomfort that is felt is largely a consequence of depth signal conflicts between accommodation and vergence, and between the associated biomechanical apparatus just described. We shall attempt to model this overall system by a second-order differential equation

$$\frac{d^2C(t)}{dt^2} + 2\zeta\omega_n\frac{dC(t)}{dt} + \omega_n^2 C(t) = k\omega_n^2 I(t) \tag{1}$$

where k, ω_n and ζ are the gain, natural frequency and damping ratio of the system, respectively. All of these values can be used to completely characterize the system, including determining the peak, settling, and rise times of the system. The transfer function of the model is then

$$H(s) = \frac{C(s)}{I(s)} = \frac{k\cdot\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}. \tag{2}$$

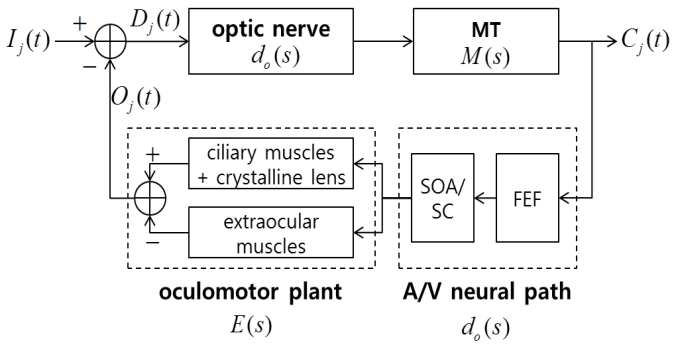


Fig. 3. Schematic representation of the proposed visual discomfort mechanism.

B. Overview of TVDM

Figure 3 is a schematic block diagram of our model (TVDM). The figure depicts passage of the space-time visual signal from the eyes, through the brain to extracortical area MT, which controls signaling via the A/V neural path, which in turn activates the ciliary muscles and extraocular muscles to control accommodation and vergence, respectively, via the oculomotor plant. Vergence velocity signals and vergence position signals reach the oculomotor plant through the oculomotor and abducens nerves. Then, the responses of the plant after the eyes are re-oriented and re-focussed serve as feedback to the system. The feedback signal is transmitted back to MT after going through the sensorimotor transformation process. Since the responses along the A/V neural path, including the optic nerve imply a delay, we denote the transfer function of these stages by $d_o(s)$ [17]. In the following, we briefly summarize the essential signals passing through the overall model. These are described in detail in the following sections.

1) Input Signal $I_j(t)$: Given a viewed S3D image j, the system input is modeled as the reciprocal of the absolute difference between the temporal accommodation and vergence signals:

$$I_j(t) = \frac{1}{|Acc_j(t) - Ver_j(t)| + \lambda} \tag{3}$$

where $Acc_j(t)$ and $Ver_j(t)$ model the aggregate accommodation and vergence responses caused by image j at time t, respectively. The perception of the S3D image by the viewer is activated at $t = 0$ and occurs over a period of T seconds. The constant λ is included to avoid instability when $Acc_j(t) \approx Ver_j(t)$ (we use $\lambda = 1$). Because the difference between accommodation and vergence is thus simply determined from the input image $I_j(t)$, it is possible to use this difference to predict depth-driven visual discomfort. $I_j(t)$ satisfies $0 \leq I_j(t) \leq 1$, where $I_j(t) = 1$ when $Acc_j(t) \approx Ver_j(t)$ (no disparity) and $I_j(t) \to 0$ as $|Acc_j(t) - Ver_j(t)| \uparrow$. Detailed explanations are described in Sec. II-C.

If disruption of the natural neural interaction between accommodation and vergence on a flat-panel stereoscopic display occurs due to excessive disparity, humans may experience an unnatural sense of depth or even annoyance, which may lead to visual discomfort.

2) Output Signal $C_j(t)$: A/V conflicts that occur when viewing an S3D image j cause levels of annoyance and/or physical discomfort that can vary over time. We model the dynamics of the overall system by

$$C_j(t) = \mathcal{L}^{-1}[d_o(s) \cdot M(s) \cdot D_j(s)] \tag{4}$$

$$= \mathcal{L}^{-1}[H(s) \cdot I_j(s)] \tag{5}$$

where $d_o(s)$ and $M(s)$ are the transfer functions of signal transmission by the optic nerve and of visual area MT, respectively. $D_j(s)$ is the Laplace transform of the input signal of the optic nerve stage as shown in Fig. 3. $C_j(t)$ should be expressed as experimentally obtained subjective assessments of visual discomfort. We obtained subjective scores recorded over the discomfort range $0 \sim 10$ using a new subjective methodology called MICSQ which allows continuous recording of human judgements, but we also normalized $C_j(t)$ to the range $0 \leq C_j(t) \leq 1$. The boundary conditions on the input and output signals are: $I_j(t) \simeq 0$, $C_j(t) \simeq 0$ (maximal visual discomfort) and $I_j(t) = 1$, $C_j(t) \simeq 1$ (no visual discomfort), respectively.

When the output signal exhibits a low value due to excessive disparity, A/V conflicts, or other factors, the eyes struggle to find a stable state and the viewer may experience visual discomfort. To model these phenomena, we regard the output of the oculomotor plant as embodying the mismatch signal between accommodation and vergence.

3) Output Signal of the Oculomotor Plant $O_j(t)$: The difference between the accommodation and vergence depth responses correlates with the energy expended by the extraocular and ciliary muscles towards attaining a stable state of comfort:

$$O_j(t) = Y_{Acc}(t) - Y_{Ver}(t) \tag{6}$$

where $Y_{Acc}(t)$ and $Y_{Ver}(t)$ are the accommodation and vergence responses. When $C_j(t) = 0$, the level of visual discomfort is large and $O_j(t)$ fluctuates considerably. Conversely, when $C_j(t) = 1$ (no visual discomfort), then $O_j(t)$ is small.

4) Input Signal to the Optic Nerve $D_j(t)$: We model the signal containing information regarding eye movements that have occurred in response to the difference between accommodation and vergence by

$$D_j(t) = \pm(I_j(t) - O_j(t)). \tag{7}$$

If $I_j(0) = 1$, then $O_j(0)$ is small and $D_j(0) \simeq 1$. $D_j(t)$ contains the feedback control signal responses of the oculomotor plant and visual area MT, As $I_j(\infty) \to O_j(\infty)$, $D_j(\infty) \downarrow 0$.

C. Accommodation, Vergence, and Foveation

1) Definition of $I_j(t)$ in (3): Accommodation is measured in *diopters* [D], which is the reciprocal of focussing depth (meters). Unlike natural viewing conditions, when viewing a stereoscopic display the crystalline lens accommodates to the display distance, while the vergence apparatus allows the

viewer to perceive depth in accordance with the compulsory disparity values. Thus, $Acc_j(t)$ is assumed constant:

$$Acc_j = \text{mean}_{(u,v)}\{Acc_j^{(u,v)}\} = \frac{1}{V}\ [D] \tag{8}$$

where the sample mean is computed over all 2D spatial coordinates (u, v), $Acc_j^{(u,v)}$ is the accommodation value at each location, and V is the viewing distance in meters $[m]$.

When a viewer first casts their gaze onto an S3D image, s/he will tend to visually scan the image content with corresponding eye movements. As the visual system gathers information and stores it in short-term sensory memory, vergence may vary independently of accommodation for about 1 second [44]. However, after this brief delay, the visual scanpaths tend to stabilize and become more regular, and the vergence and accommodation processes engage each other. This is particularly true when viewing an S3D display, where large head movements are unusual [17]. This can be confirmed from recorded visual scanpaths using an eye-tracker, as shown later in Sec. IV-D (Fig. 10).

Vergence can be expressed in terms of *meter angles* [*MA*], the reciprocal of fixation distance (also meters). Since the units of accommodation and vergence are the same, the accommodation and vergence responses can be readily compared. $Ver_j(t)$ in (3) is also determined by tracking the average value, $Ver_j = \text{mean}_{(u,v)}\{Ver_j^{(u,v)}\}$. Then,

$$I_j(t) \simeq I_j = \frac{1}{|Acc_j - Ver_j| + \lambda}. \tag{9}$$

2) Accounting for Saliency and Foveation: Under natural viewing conditions, the binocular visual system coordinates the movement of the two eyes, so that both eyes fixate at the same point whose images are projected onto the foveas [31]. The approximation $Ver_j = \text{mean}_{(u,v)}\{Ver_j^{(u,v)}\}$ does not hold in most cases for the following reasons. First, the retinal photoreceptors are non-uniformly distributed with highest density at the fovea; the density decreases reciprocally with distance from the fovea [31], [32]. Second, vergence is also affected by visual attraction (saliency). For example, some sequences may produce visual discomfort over only a small percentage of the visual field. In such a case, a viewer may tend to direct the point of gaze away from such regions, causing them to be sampled at lower (non-foveal) resolution. Ideally, the design of a reliable vergence response Ver_j should account for both foveation and saliency.

3) Foveation and Saliency Strength: The acuity of the HVS is a function of spatial photoreceptor density. A commonly used foveation model that fits human data well was presented in [33]. The foveation model is expressed

$$f_c\{V,(u,v)\} = \min\left[\frac{e_2 \ln(\frac{1}{CT_0})}{\alpha\{e_2 + \arctan(\frac{d(x)}{VW})\}}, \frac{\pi V W}{360}\right] \tag{10}$$

where $d(x) = \sqrt{(\frac{u-u_f}{m})^2 + (\frac{v-v_f}{m})^2}$ is the distance (in pixels) between point (u, v) and the fixation point (u_f, v_f), m is a screen magnification factor, $\arctan(\frac{d(x)}{VW})$ is the eccentricity, W is the display width and V is the assumed viewing distance. Geisler and Perry [33] found values for the spatial frequency decay constant $\alpha = 0.106$, half-resolution eccentricity constant $e_2 = 2.3$, and minimum contrast threshold $CT_0 = 1/64$.

We follow the hypothesis that human visual fixations are drawn to salient 2D or 3D image regions [34], [35]. Here, we assume that a region having high saliency falls on the fovea and hence is captured at highest resolution. The 3D saliency prediction model in [35] computes saliency strength $\mathcal{S}_j^{(u,v)}$ in image j by incorporating luminance, disparity, and visual comfort factors. Figure 4 depicts examples of saliency strength maps computed on images from the IEEE-SA [36] and EPFL [37] S3D image databases. The middle column in Fig. 4 shows exemplar saliency strength weight maps, while the right column shows recorded visual fixations using an eye-tracker. The saliency strength maps (Fig. 4 (b), (f), (j) and (n)) clearly are correlated with the subjective results obtained using the eye-tracker.

4) TVDM Input Signal: The vergence signal model for an S3D image j incorporates both saliency strength and foveation:

$$\begin{aligned} Ver_j(t) &\simeq Ver_j \\ &= \text{mean}_{(u,v)}\{\mathcal{S}_j^{(u,v)} \cdot f_c\{V,(u,v)\} \cdot Ver_j^{(u,v)}\}. \end{aligned} \tag{11}$$

Using (11), one can construct a virtual view of an image weighted by foveation and saliency strength. Hence, (3) becomes

$$\begin{aligned} I_j(t) &\simeq I_j \\ &= \frac{1}{|\frac{1}{V} - \text{mean}_{(u,v)}\{\mathcal{S}_j^{(u,v)} \cdot f_c\{V,(u,v)\} \cdot Ver_j^{(u,v)}\}| + \lambda}. \end{aligned} \tag{12}$$

III. TVDM System Identification

In order to be able to better understand the temporal aspects of visual discomfort levels felt by humans viewing S3D images over time, we conducted a subjective study on 350 stereo images drawn from the IEEE-SA, EPFL and DIML [12] databases using the new, multimodal subjective study methodology called MICSQ [19]. We used the collected subjective data to study the relationship between quantitative subjective judgements, $C_j(t)$ for each S3D image j and corresponding input I_j. Based on this, we develop a transfer function model using non-linear regression.

A. Transfer Function of TVDM

As stated in Sec. I-A, TVDM models the function of neural pathways and processing elements between the eyes and certain areas of the brain by a second-order system with transfer function $H(s) = k \cdot \omega_n^2/(s^2 + 2\zeta\omega_n s + \omega_n^2)$. In the following, we show that the interaction between brain activity in visual area MT and the corresponding extraocular muscles and lens dynamics expressed by TVDM transfer function that can be used to predict the visual discomfort levels experienced when viewing S3D images.

First, the A/V neural path including all nerves are modeled as simple delay elements having a transfer function e^{-Ts},

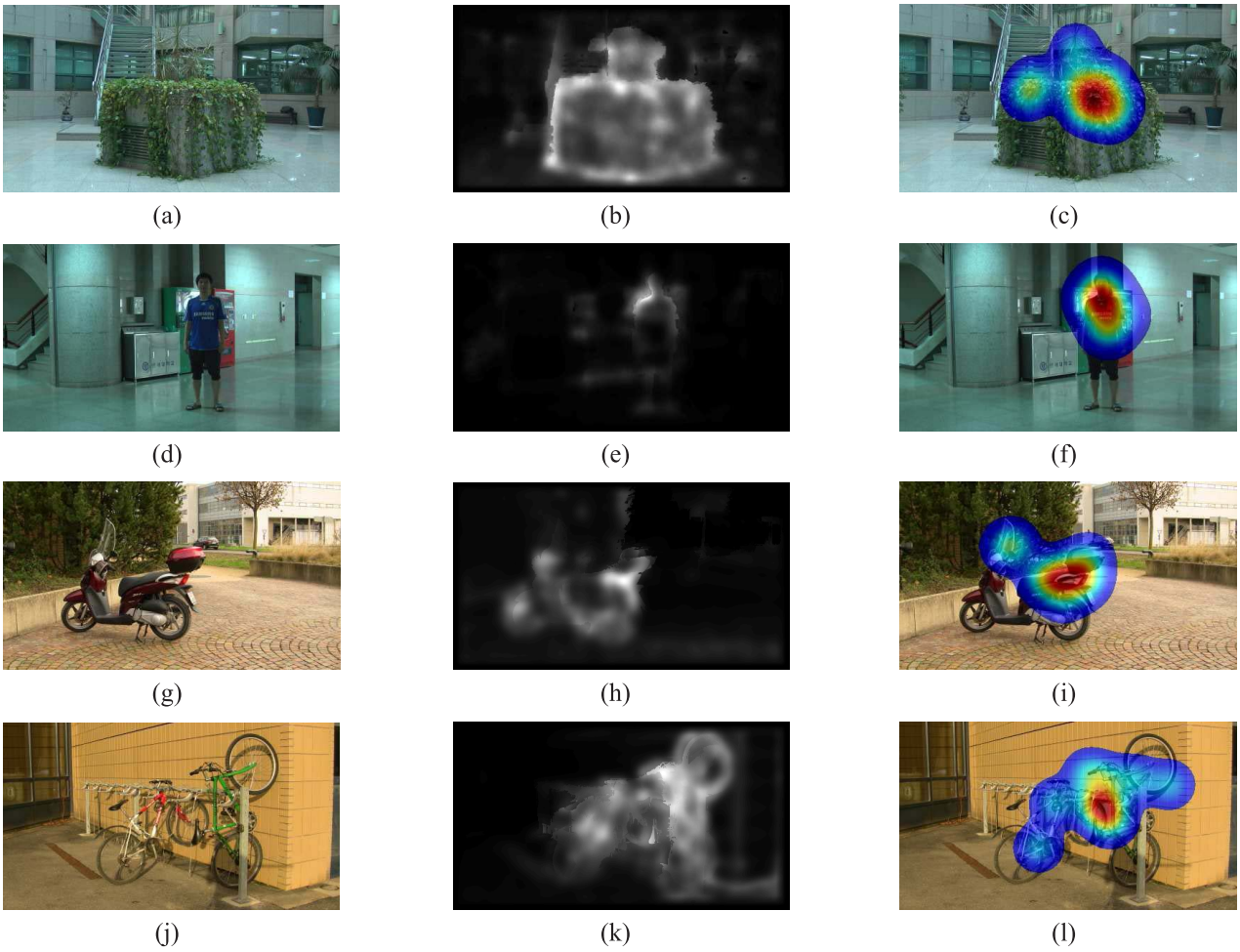


Fig. 4. Computed saliency strength map and map of recorded visual fixations using an eye-tracker. Left column: original left images (a) "ISL5-50" (d) "ISL20-100" from the IEEE-SA database [36] and (g) "07-01" (j) "09-01" from the EPFL database [37]. Middle column: (b), (e), (h) and (k) computed saliency strength map. Right column: results of eye-tracker simulation ((c), (f), (i) and (l)).

where T is the delay time [16], [17]. Linearizing this by Taylor approximation yields the following approximate transfer function model of the optic and A/V neural path stages

$$d_o(s) = 1 - Ts \tag{13}$$

where $T = 0.05$ seconds as assumed in [17].

Second, the model of the transfer function of the oculomotor plant

$$E(s) = \frac{O_j(s)}{N(s)} = H_A(s) - H_V(s) \tag{14}$$

where $N(s)$ is the Laplace transform of the output of the A/V neural path stage, $H_A(s) = \frac{0.25}{0.25s+1}$ and $H_V(s) = \frac{0.3}{0.15s+1}$, as depicted in Fig. 5 [15].

Finally, the transfer function of relevant control signal apparatus in visual area MT, denoted $M(s)$, represents the time-variant neural activity between the input signal carried through the optic nerve and the output signal transmitted along the A/V neural path. The dynamic behavior of visual area MT has been derived from brain activity measurements in a number of biological and physiological studies [23], [39].

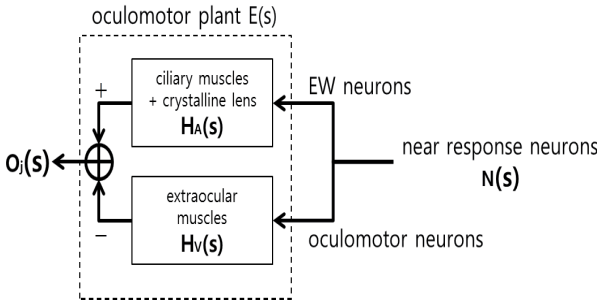


Fig. 5. Input and output signal flows of the oculomotor plant $E(s) = O_j(s)/N(s)$. To achieve accommodation, the EW neurons of near response neurons control the ciliary muscles and the crystalline lens. To cause vergence, the oculomotor neurons control the extraocular muscles. The plants representing the accommodation and vergence processing units denoted by transfer functions $H_A(s)$ and $H_V(s)$, respectively.

Since the transfer function $H(s)$ in Fig. 3 is

$$H(s) = \frac{k\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} = \frac{d_o(s)M(s)}{1 + d_o^2(s)E(s)M(s)}, \tag{15}$$

the transfer function of visual area MT is modelled as

$$M(s) = \frac{k\omega_n^2}{d_o(s)\{s^2 + 2\zeta\omega_n s + \omega_n^2(1 - kd_o(s)E(s))\}}. \tag{16}$$

B. Regression Analysis With MICSQ

1) Assessment Environment: We conducted a fairly large-scale human study of visual discomfort felt when viewing S3D images using the MICSQ subjective study method [19]. MICSQ uses a multimodal (aural and tactile) interactive and continuous (if needed) scoring process which helps engage and focus subjects on their tasks. Several 3D experiments were conducted on a 46-inch polarized stereoscopic display having resolution 1920×1080 and display height 0.6 m, so the viewing distance for each subject was set to about 1.8 m, following ITU-R Rec. 500-13 [13]. The study was conducted under constant room and background illumination conditions. Interactive responses were recorded using a tablet-PC (model: Samsung Galaxy-Tab (SEC-SHW M180W)).

2) Subjects: Thirty nine subjects (27 male and 12 female) participated in the subjective study. Five of the subjects were previously involved in 3D research, while the others were inexperienced. The ages of the subjects ranged from 24 to 31 years with an average age of 27. All subjects were tested and found to have good or corrected visual acuity of greater than 1.25 (using the Landolt C-test) and good stereoscopic acuity of less than 60 arc (the RANDOT stereo test). Each subject was asked to assess the degree of visual (dis)comfort experienced when viewing each S3D image for 10 seconds. The range of subjective scores was set from $0 \sim 1$ with an initial score of 0 and with equally spaced marks on a Likert scale [bad]-[poor]-[fair]-[good]-[excellent], following ITU-R Rec. 500-13. Subjective scores were continuously collected over each 10-second presentation, at a sampling rate of 10 Hz. If a subject's rating score was very different from the group results, the subject was regarded as an outlier following the rejection procedure in [13]. However, no outlier was detected during the course of the statistical analysis. In addition, to enable the subjects to familiarize themselves with MICSQ and the evaluation process, ten S3D sample images were presented to each subject prior to commencing the subjective task. Each subject's sessions were separated by a one week interval to eliminate any residual visual fatigue.

3) Regression: The transfer functions $H(s)$ and $M(s)$ were estimated using non-linear least squares regression (the Matlab function "nlift") to determine the optimal coefficients. A total of 350 high-resolution (1920×1080) S3D images were used: 220 stereo images from the IEEE-SA database, 90 stereo images from the EPFL database, and 40 stereo images from the DIML database (indexed $j = 1, \ldots, 350$). We develop the relationship between the input $I_j(t)$ and predicted visual discomfort level $C_j(t)$ in Laplace space: from (2),

$$C_j(s) = H(s) \cdot I_j(s) = \frac{k\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \cdot \frac{I_j}{s} \tag{17}$$

which yields the temporal prediction

$$C_j(t) = kI_j \cdot \left(1 - \frac{1}{\sqrt{1-\zeta^2}} e^{-\zeta\omega_n t} \cos(\omega_n \sqrt{1-\zeta^2} t - \phi)\right) \tag{18}$$

where $\phi = \arctan(\zeta/\sqrt{1-\zeta^2})$.

In order to obtain the coefficients k, ω_n and ζ via regression, divide (18) by I_j: $C_j(t)/I_j = k \cdot \left(1 - \frac{1}{\sqrt{1-\zeta^2}} e^{-\zeta\omega_n t} \cos(\omega_n \sqrt{1-\zeta^2} t - \phi)\right)$. The best fit was found by minimizing the least squares error between the vector of mean opinion scores (MOSs) $[\mathrm{MOS}_j;\ 1 \le j \le 350]$ and the vector $[C_j(t)/I_j;\ j = 1, 2, \ldots, 350]$ on the time axis ($0 \le t \le 10$ [sec]). Because the subjective assessment was conducted by multiple subjects, MOS is computed as

$$\mathrm{MOS}_j(t) = \frac{\sum_{n=1}^{N} C_j(t)}{N} \tag{19}$$

where $N = 39$ is the number of subjects. The stereo images from IEEE-SA, EPFL, and DIML were separated into 80% and 20% subsets for regression and test, respectively. In other words, 280 different continuous MOSs obtained from thirty nine subjects were fit to (18), then the correlation with the remaining 70 subjective results were calculated for each trial. To ensure that our results were not affected by a specific regression-test separation, we repeated the regression 2000 times using randomly selected 80% - 20% subsets at each trial.

The regression yielded $k = 4.27$, $\zeta = 0.73$ and $\omega_n = 0.75$. Hence, the overall transfer function model used in TVDM is expressed as

$$H(s) = \frac{2.402}{s^2 + 1.095s + 0.563}. \tag{20}$$

In addition, the transfer function model of the relevant processing and signalling elements of visual area MT becomes

$$M(s) = \frac{2.402}{(1-0.05s)\{s^2 + 1.095s + 0.563 - (4.27 - 0.214s)E(s)\}} \tag{21}$$

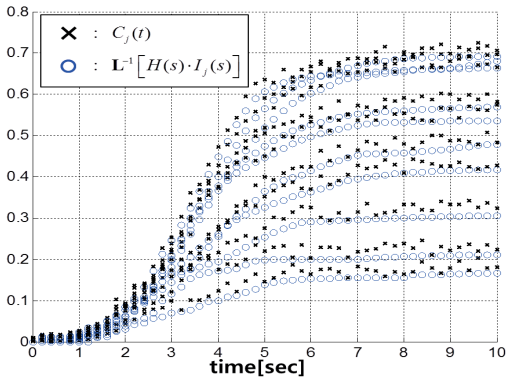
where $E(s)$ is the transfer function of the oculomotor plant.

IV. Results and Analysis of TVDM

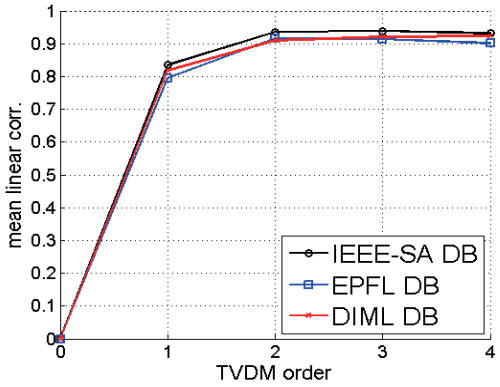
A. Statistical Performance Evaluation

To evaluate the accuracy of the regression process, we measured the linear correlation coefficient between the two time series $C_j(t)$ and $\mathcal{L}^{-1}[H(s) \cdot I_j(s)]$. Figure 6 (a) depicts plots of these two time series for 70 arbitrary S3D images. Very high linear correlation coefficients of 0.936 and 0.917 were obtained on the IEEE-SA and EPFL database images, respectively. The graphs of $\mathcal{L}^{-1}[H(s) \cdot I_j(s)]$ accurately matched the corresponding subjective traces $C_j(t)$ as shown in this figure. In addition, to demonstrate that the regression process does not heavily depend on the TVDM order, we measured the mean linear correlation coefficients as a function of TVDM order, as plotted in Fig. 6 (b). The value of correlation increases with the TVDM order, but only very slightly above order 2.

To obtain a visual discomfort score for each image, TVDM was measured score at $t = 10$ seconds. Figure 7 plots the 10-second TVDM score for each image ($j = 1, 2, \ldots, 350$) versus MOS. The results qualitatively demonstrate that TVDM scores correlate highly with MOSs.



(a)



(b)

Fig. 6. (a) Best least-squares fits between the time series $C_j(t)$ and $\mathcal{L}^{-1}[H(s) \cdot I_j(s)]$ for 70 arbitrary S3D images of IEEE-SA database. (b) Mean linear correlation coefficients of TVDM against MOS as a function of model order using the IEEE-SA, EPFL, and DIML databases.

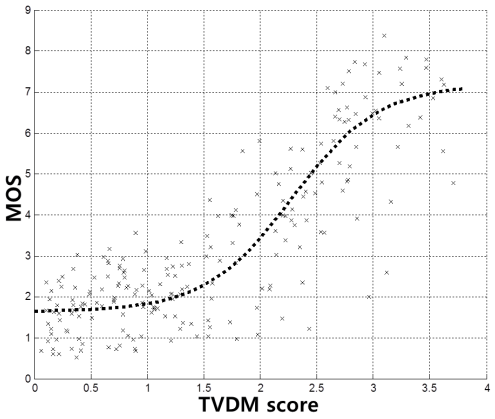


Fig. 7. Scatter plot of TVDM at $t = 10$ [sec] vs. MOS on the IEEE-SA and EPFL 3D database images.

The mean values of the Pearson linear correlation coefficient (LCC) and the Spearman rank order correlation coefficient (SROCC) were 0.8904 and 0.8905, respectively.

We compared the performance of TVDM with those of prior objective models proposed by Yano *et al.* [8], Kim and Sohn [12], Nojiri *et al.* [40], Choi *et al.* [41] and Park *et al.* [1]. Yano *et al.* [8] computed the ratio of sums of disparities near the screen to those far from the screen, by assuming a comfort zone spanning 60 *arcmin*.

TABLE I

LCC AND SROCC FOR 2000 TRIALS OF RANDOMLY CHOSEN REGRESSION AND TEST SETS ON THE IEEE-SA DATABASE

	LCC			SROCC		
	Mean	**Median**	**Std.**	**Mean**	**Median**	**Std.**
Yano [8]	0.4013	0.4274	0.0728	0.3965	0.3978	0.0712
Kim [12]	0.7412	0.7333	0.0721	0.6895	0.6998	0.0640
Nojiri [40]	0.7104	0.7095	0.0588	0.7302	0.6524	0.0612
Choi [41]	0.7209	0.7023	0.0624	0.7133	0.7049	0.0658
Park [1]	0.8604	0.8672	0.0482	0.7831	0.7882	0.0451
TVDM	**0.8931**	**0.8954**	**0.0405**	**0.8973**	**0.8815**	**0.0422**

TABLE II

LCC AND SROCC FOR 2000 TRIALS OF RANDOMLY CHOSEN REGRESSION AND TEST SETS ON THE EPFL DATABASE

	LCC			SROCC		
	Mean	**Median**	**Std.**	**Mean**	**Median**	**Std.**
Yano [8]	0.8471	0.8536	0.0515	0.8721	0.8732	0.0604
Kim [12]	0.8593	0.8451	0.0632	0.8852	0.8763	0.0614
Nojiri [40]	0.8415	0.8532	0.0591	0.8633	0.8598	0.0583
Choi [41]	0.7539	0.7928	0.0631	0.8053	0.8381	0.0603
Park [1]	0.8984	0.8817	0.0532	0.9045	0.8987	0.0585
TVDM	**0.8995**	**0.8893**	**0.0505**	**0.9108**	**0.9063**	**0.0532**

TABLE III

LCC AND SROCC FOR 2000 TRIALS OF RANDOMLY CHOSEN REGRESSION AND TEST SETS ON THE DIML DATABASE

	LCC			SROCC		
	Mean	**Median**	**Std.**	**Mean**	**Median**	**Std.**
Yano [8]	0.7931	0.7452	0.0631	0.6732	0.7053	0.0611
Kim [12]	0.8645	0.8595	0.0612	0.8994	0.8836	0.0605
Nojiri [40]	0.8231	0.8184	0.0544	0.8152	0.8139	0.0563
Choi [41]	0.8043	0.8028	0.0553	0.8209	0.8322	0.0635
Park [1]	0.8511	0.8673	0.0499	0.8533	0.8617	0.0502
TVDM	**0.8827**	**0.8786**	**0.0443**	**0.8635**	**0.8669**	**0.0493**

Kim and Sohn [12] proposed several features to predict the visual discomfort experienced when viewing S3D images, most notably the disparity range and the maximum angular disparity, based on an assumed comfortable viewing zone. Nojiri *et al.* [40] used a number of sample statistics descriptive of experienced parallax, such as minimum and maximum disparity, disparity range, and the dispersion and absolute average of disparity. Choi *et al.* [41] proposed three kinds of features descriptive of the spatial, temporal, and differential elements of viewing stereoscopic 3D video. Park *et al.* [1] extracted four kinds of features descriptive of the anomaly of vergence-accommodation/vergence (VA/V) ratio, anomaly of accommodation-vergence/accommodation (AV/A) ratio, absence of de-focus blur and absence of differential blur in S3D images.

Tables I, II, and III show the mean, median, and standard deviation values of LCC and SROCC for these prior models over the 2000 regression and test trials on the IEEE-SA, EPFL, and DIML database images, respectively. The correlation values from the previous algorithms were obtained using a

TABLE IV
RESULTS OF F-TEST PERFORMED ON THE RESIDUALS BETWEEN OBJECTIVE MODELS INCLUDING TVDM AND MOS

	Yano [8]	Kim [12]	Nojiri [40]	Choi [41]	Park [1]	**TVDM**
Yano [8]	-	0	-	0	0	0
Kim [12]	1	-	1	1	0	0
Nojiri [40]	-	0	-	0	0	0
Choi [41]	1	0	1	-	0	0
Park [1]	1	1	1	1	-	0
TVDM	1	1	1	1	1	-

support vector regression method with the same number of trials [42], [43]. As shown in both tables, TVDM delivered the highest LCC and SROCC values and the minimum standard deviation values by a substantial margin.

In addition, to investigate the statistical significance of the relative performances of the prior objective models that were tested, we conducted standard F-tests on the errors between MOS and the objective model results. The residual error between the prediction score of each objective model and MOS on the IEEE-SA, EPFL and DIML database images was used to evaluate the possible statistical superiority of one objective model against another. The residual errors between the prediction score of an objective model and MOS is given by:

$$\text{Model Residual} = \{S_j - \text{MOS}_j, j = 1, \ldots, 350\} \quad (22)$$

where S_j is the vector of fitted objective scores for stereo image j. As shown in Table IV, F-tests were conducted on the ratios of the variance (22) for each objective model relative to every other one at the 95% significance level [38]. A value of '1' indicates that the statistical performance of the objective model in the row is superior to that of the model in the column and a symbol value of '0' indicates that the objective model in the row is statistically inferior to that of the model in the column. A symbol of '-' indicates that the statistical performance of the model in the row is equivalent to that of the model in the column.

B. Response Time

The response time of the visual discomfort prediction model is defined as the time required to reach a final predicted subjective score. This depends on a variety of factors. For example, if the viewer does not perceive that the input image "looks like" a real-world scene, more response time may be needed to reach a final decision, as observed in [44]. We interpreted the settling time (T_s) of the second-order system to be the time required for the output to reach and stay at 2% of its final value. Therefore, the response time needed to compute a visual discomfort score on an S3D image can be viewed as a simple linear function of the settling time of TVDM. Since the settling time is defined by the time at which the envelope decays to 0.02; $\frac{1}{\sqrt{1-\zeta^2}} e^{-\zeta\omega_n t} = 0.02$, the settling time is obtained as

$$T_s = \frac{-\ln(0.02\sqrt{1-\zeta^2})}{\zeta\omega_n}. \quad (23)$$

TABLE V
MEAN VALUE OF $I_{j\in g}$ AND CORRESPONDING SETTLING TIME T_s FOR EACH GROUP

Group	Group 1	Group 2	Group 3	Group 4	Group 5
mean$[I_{j\in g}]$	0.9581	0.8637	0.7093	0.5125	0.3817
T_s	4.4938	5.1329	5.8060	10.0201	23.1946

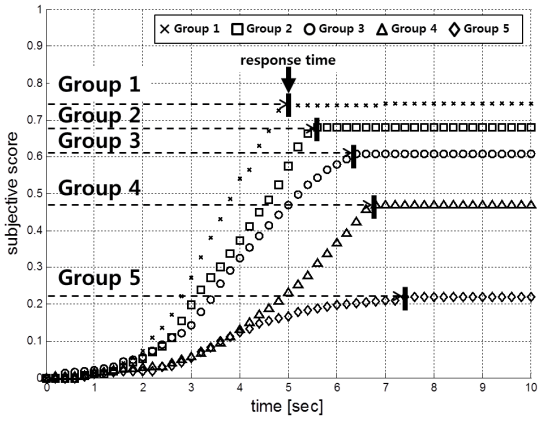


Fig. 8. Mean subjective temporal score variations among the 5 groups.

TABLE VI
MEAN VALUE OF $I_{j\in g}$ AND CORRESPONDING GAIN k FOR EACH GROUP

Group	Group 1	Group 2	Group 3	Group 4	Group 5
mean$[I_{j\in g}]$	0.9581	0.8637	0.7093	0.5125	0.3817
$k_{j\in g}$	2.8670	3.5591	4.9372	6.0648	6.8527

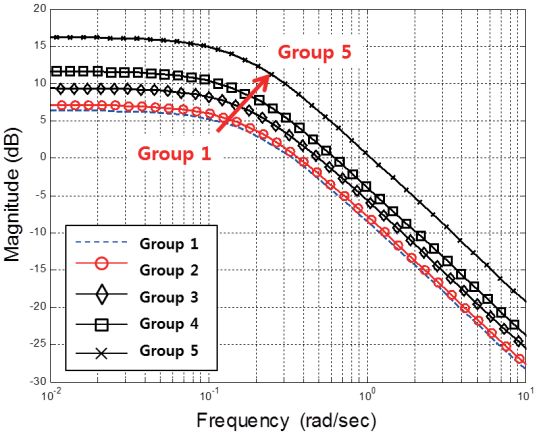


Fig. 9. Bode magnitude curves of the 5 groups against frequency.

In order to investigate the relationship between response time and the difference between the accommodation and vergence signals, we divided the 350 S3D images used into five groups depending on their values I_j. Both databases consist of five different pairs of stereo images using multiple evenly separated convergence points and separating the baseline distance, respectively. Therefore, all test images

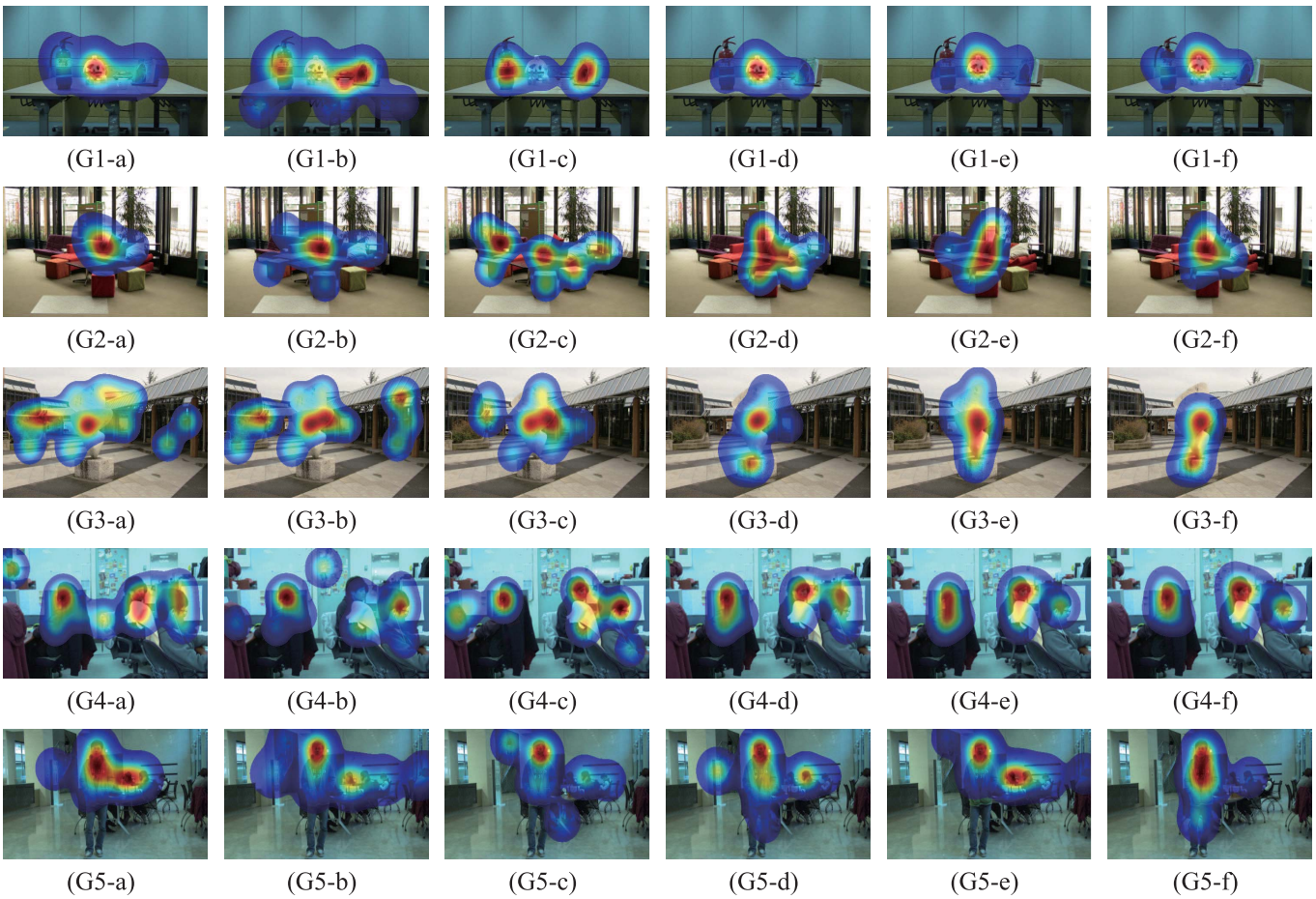


Fig. 10. Eye-tracker (60 Hz) results on (G1) "ISS4$_0$" for Group 1, (G2) "03-01" for Group 2, (G3) "05-07" for Group 3, (G4) "INS1$_{100}$" for Group 4, (G5) "ISL14$_0$" for Group 5 at (G#-a) 0.25, (G-b) 0.5, (G-c) 0.75, (G-d) 1, (G-e) 5, and (G-f) 10 seconds.

from both databases have diverse depth statistics. In each group, 70 images have similar values of I_j with a group index, g (Group 1∼Group 5) in order of mean values $I_{j\in g}$. Table V shows the mean values, $I_{j\in g}$ of each group where $I_{j\in Group\ i+1} > I_{j\in Group\ i}$. This means that a viewer feels more comfortable when viewing S3D images from lower indexed groups.

Figure 8 depicts the mean temporal subjective scores for each group. Broadly, the response time required to determine a visual discomfort score increases with the mean value $I_{j\in g}$. In other words, when viewing a comfortable S3D image, the subjects tend to determine subjective visual discomfort scores more quickly than when they are watching an uncomfortable S3D scene. One possible explanation for this is scene familiarity. If an S3D scene is familiar to a viewer, (s)he tends to determine a degree of visual discomfort more quickly [44].

We also obtained the settling time, T_s, for each group g using the regression technique on the 70 S3D images in each group. As revealed by Table V, the value of T_s tended to increase with decreases in the mean value of $I_{j\in g}$.

C. System Gain

By the final value theorem, the gain of TVDM is $\lim_{s\to 0} H_j(s) = k$. Whereas the final subjective visual discomfort score at $t = 10$ seconds is

$$C_j(10) = \lim_{s\to 0} H_j(s)\cdot I_j, \\ = kI_j\cdot\left(1-\frac{1}{\sqrt{1-\zeta^2}}e^{-10\zeta\omega_n} \times\cos(10\omega_n\sqrt{1-\zeta^2}-\phi)\right) \quad (24)$$

where $\phi = \arctan(\zeta/\sqrt{1-\zeta^2})$. The gain indicates the perceptual effort exerted by TVDM to predict the level of visual discomfort experienced when viewing an S3D image. Since variations in the gain greatly depend on the difference between accommodation and vergence, we determined the values of k for each group g using the regression process. As shown in Table VI, the mean value of k was highest for Group 5 as compared with the other groups, meaning that greater perceptual effort was required as the difference between accommodation and vergence increased. This is also observed using Bode plots [45]. The Bode magnitude plot has a linear decibel [dB] scale for values on the vertical axis and a logarithmic frequency scale on the horizontal axis. The Bode phase plot represents the system phase as a function of frequency.

As shown by the arrow in Fig. 9, as the group number increases, the gain increases and the gap between adjacent

groups widens. This means that a large perceptual effort needs to be made when the input stimulus causes visual discomfort when being viewed. When a subject views a comfortable S3D image, little effort is required to perceive 3D scene information.

D. Gaze Shifting and Relative Disparity

In order to gather and interpret visual information, viewers tend to scan the display screen when they commence viewing S3D image. The informations gathered are processed via multiple neural pathways, the responses to which drive the biomechanic visual apparatus implicated in the experience of visual discomfort. This can produce time-varying muscular stresses in the oculomotor response. For example, if the differences in disparity of the gaze points vary significantly, then the subjective experience of (dis)comfort may considerably fluctuate around a level of poor comfort [44]. Likewise, the unit step response graphs of the general second-order system exhibit varying degrees of oscillation and overshoot depending on the damping ratio ζ. The tendency is that strong oscillations are driven by reduced values of ζ. Thus, it is of interest to investigate the effects of relative disparity as a function of scanpaths on variations in experienced visual discomfort by measurement of the TVDM damping ratio.

Figure 10 shows eye-tracker results on five representative S3D images from each group recorded during the subjective visual discomfort assessment experiments. The visual fixation maps recorded at time instants at 0.25, 0.5, 0.75, 1, 5, and 10 seconds from beginning of presentation are presented by column in Fig. 10 starting from the left. It may be observed that gaze shifts occurred frequently until about one second passed, with considerably reduced frequency afterwards. In addition, as may be seen in the leftmost columns of Fig. 10, visual attention tended towards the center of the screen, a well-known center bias phenomena [35].

In order to investigate the relationship between the relative disparity variations occurring at the sequences of fixation points and variations in the subjective scores, we calculated the damping ratio for each group g using a regression process. Moreover, to quantify variations in disparity amongst sequences of fixations, we computed the standard deviation of disparity at all fixation points over five second intervals:

$$\mathcal{R} = \sqrt{\frac{1}{N}\sum_{n}(d(n)-\mu)^2} \tag{25}$$

where $\mu = \frac{1}{N}\sum_n d(n)$, $d(n)$ is angular disparity at fixations, and N is the total number of disparity measurements over the interval.[2] Since there is a certain tendency that the greater is the dispersion of disparities among a sequence of fixation points, the greater the likelihood that A/V conflicts will occur leading to consequent visual discomfort [1].

As shown in Table VII, the mean value of ζ was lowest for Group 5 amongst all groups, indicating a greater degree of disparity oscillation. It may be observed that the subjective scores tend to fluctuate in accordance with relative disparity variations.

[2]We calculated the standard deviation over non-overlapping 5 second intervals of eye-tracking.

TABLE VII
MEAN VALUE OF $I_{j\in g}$ AND CORRESPONDING DAMPING RATIO ζ FOR EACH GROUP

Group	Group 1	Group 2	Group 3	Group 4	Group 5
mean$[I_{j\in g}]$	0.9581	0.8637	0.7093	0.5125	0.3817
$\mathcal{R}$	0.001	0.006	0.011	0.023	0.057
$\zeta_{j\in g}$	0.952	0.863	0.782	0.538	0.326

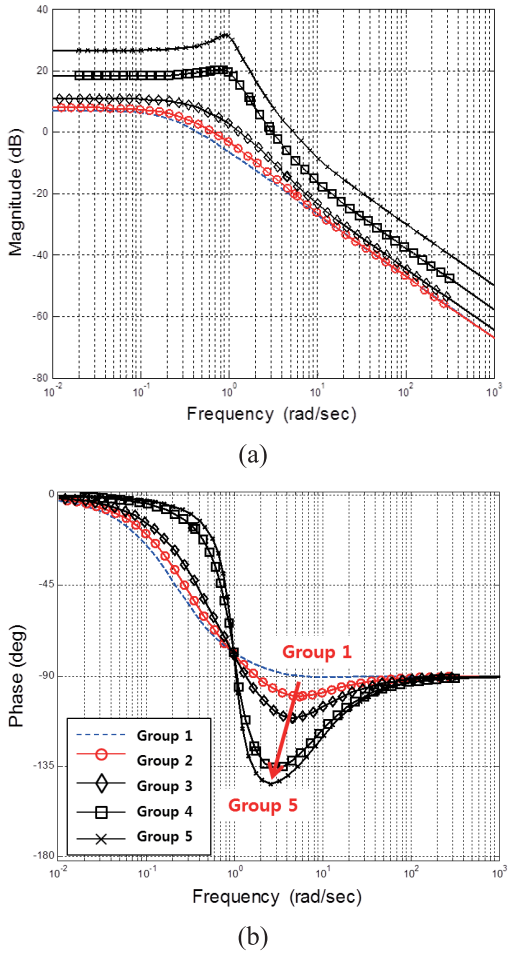


Fig. 11. Bode (a) magnitude and (b) phase plots of visual area MT for each of the 5 groups.

E. Discussion of Visual Area MT

Using the results of the regression trials, we obtained the transfer function model of relevant processing elements of visual area MT, $M(s)$, using the transfer function of the oculomotor plant, as shown in Eq. (21). The transfer function has 15 poles and 13 zeros; the model frequency response of visual area MT exhibits low-pass filter characteristics.

The processing of visual data along the dorsal stream of visual information by visual area MT is associated with the

computation of motion, disparity, and location, as well as supplying signaling that drives vergence eye movements and the accommodation response. Previous magnetoencephalography (MEG) studies suggest that the temporal responses of these are tuned to low frequencies [46], [47]. Lutzenberger *et al.* [48] found that the peak amplitude of the visual-evoked potential (VEP) to lie at approximately 10 Hz in adults using electroencephalography. Moreover, in [39], activation in visual cortex was found to peak within a low temporal frequency band using positron emission tomography (PET).

Next, in order to determine the model frequency response of the relevant components of visual area MT as a function of the difference between accommodation and vergence, we studied the Bode plots of $M(s)$ for each group g as shown in Figs. 11 (a) and (b). The Bode magnitude curve for Group 5 occupies a slightly higher frequency band than the other groups, as shown in Fig. 11 (a). Prior results suggest that high frequencies are more easily observed in stressful or uncomfortable situations [49], [50]. Although the differences between the results for Group 1 and Group 2 are not significant, the Bode magnitude graph of Group 1 does occupy the lowest frequency band. Conversely, the frequency responses of Groups 4 and 5 occupy higher frequency bands, exhibit higher gain and have noticeable overshoot, as shown in the Bode magnitude plot of Fig. 11 (a). Further, the Bode phase curve falls below $-135°$, as shown in Fig. 11 (b).

V. Conclusions

We have described a new model and computational approach to the temporal prediction of the degree of visual discomfort experienced when viewing an S3D image based on a simple transfer function model of physiological mechanisms involved in visual accommodation and vergence. The transfer function model can be used to accurately predict the degree of visual discomfort felt. We also studied the stability of the model as it relates to predicted differences between accommodation and vergence. As advanced techniques for 3D display and demand for 3D content continue to expand, we believe that models of this type will play an important role in understanding and predicting the neural basis of visual discomfort.

References

[1] J. Park, S. Lee, and A. C. Bovik, "3D visual discomfort prediction: Vergence, foveation, and the physiological optics of accommodation," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 3, pp. 415–427, Jun. 2014.

[2] M. Lambooij, W. IJsselsteijn, M. Fortuin, and I. Heynderickx, "Visual discomfort and visual fatigue of stereoscopic displays: A review," *J. Imag. Sci. Technol.*, vol. 53, no. 3, p. 030201, 2009.

[3] W. J. Tam, F. Speranza, S. Yano, K. Shimono, and H. Ono, "Stereoscopic 3D-TV: Visual comfort," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 335–346, Jun. 2011.

[4] M. Lambooij, W. IJsselsteijn, D. G. Bouwhuis, and I. Heynderickx, "Evaluation of stereoscopic images: Beyond 2D quality," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 432–444, Jun. 2011.

[5] F. Zilly, J. Kluger, and P. Kauff, "Production rules for stereo acquisition," *Proc. IEEE*, vol. 99, no. 4, pp. 590–606, Apr. 2011.

[6] J. DeFilippis, "3D sports production at the London 2012 olympics," *SMPTE Motion Imag. J.*, vol. 122, no. 1, pp. 20–23, Jan./Feb. 2013.

[7] S. Yano, M. Emoto, and T. Mitsuhashi, "Two factors in visual fatigue caused by stereoscopic HDTV images," *Displays*, vol. 25, no. 4, pp. 141–150, 2004.

[8] S. Yano, S. Ide, T. Mitsuhashi, and H. Thwaites, "A study of visual fatigue and visual comfort for 3D HDTV/HDTV images," *Displays*, vol. 23, no. 4, pp. 191–201, 2002.

[9] F. Speranza, W. J. Tam, R. Renaud, and N. Hur, "Effect of disparity and motion on visual comfort of stereoscopic images," in *Proc. Stereosc. Displays Virt. Real. Syst. XIII*, vol. 6055. 2006, pp. 60550B-1–60550B-9.

[10] T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks, "The zone of comfort: Predicting visual discomfort with stereo displays," *J. Vis.*, vol. 11, no. 8, pp. 1–29, 2011.

[11] I. Ohzawa, G. C. Deangelis, and R. D. Freeman, "Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors," *Science*, vol. 249, no. 4972, pp. 1037–1041, 1990.

[12] D. Kim and K. Sohn, "Visual fatigue prediction for stereoscopic image," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 2, pp. 231–236, Feb. 2011.

[13] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document ITU-Rec. BT.500-13, 2012.

[14] M. S. Shiners, *A Guide to Systems Engineering and Management*. Lexington, MA, USA: Lexington Books, 1976.

[15] C. M. Schor, "A dynamic model of cross-coupling between accommodation and convergence: Simulations of step and frequency responses," *Optometry Vis. Sci.*, vol. 69, no. 4, pp. 258–269, Apr. 1992.

[16] A. S. Eadie and P. J. Carlin, "Evolution of control system models of ocular accommodation, vergence and their interaction," *Med. Biol. Eng. Comput.*, vol. 33, no. 4, pp. 517–524, 1995.

[17] L. Peretto, E. Pivello, R. Tinarelli, and A. E. Emanuel, "Theoretical analysis of the physiologic mechanism of luminous variation in eye-brain system," *IEEE Trans. Instrum. Meas.*, vol. 56, no. 1, pp. 164–170, Feb. 2007.

[18] B. Gittleman, T. E. Dwan, and C. S. Smiley, "System identification: Human tracking response," *IEEE Trans. Educ.*, vol. 35, no. 1, pp. 31–37, Feb. 1992.

[19] T. Kim, J. Kang, S. Lee, and A. C. Bovik, "Multimodal interactive continuous scoring of subjective 3D video quality of experience," *IEEE Trans. Multimedia*, vol. 16, no. 2, pp. 387–402, Feb. 2014.

[20] H. R. Wilson and J. R. Bergen, "A four mechanism model for threshold spatial vision," *Vis. Res.*, vol. 19, no. 1, pp. 19–32, 1979.

[21] D. W. Dong and J. J. Atick, "Temporal decorrelation: A theory of lagged and nonlagged responses in the lateral geniculate nucleus," *Netw., Comput. Neural Syst.*, vol. 6, no. 2, pp. 159–178, Feb. 1995.

[22] L. M. Martinez and J.-M. Alonso, "Complex receptive fields in primary visual cortex," *Neuroscientist*, vol. 9, no. 5, pp. 317–331, Oct. 2003.

[23] A. W. Roe, A. J. Parker, R. T. Born, and G. C. DeAngelis, "Disparity channels in early vision," *J. Neurosci.*, vol. 27, no. 44, pp. 11820–11831, Oct. 2007.

[24] G. C. DeAngelis and T. Uka, "Coding of horizontal disparity and velocity by MT neurons in the alert macaque," *J. Neurophysiol.*, vol. 89, no. 2, pp. 1094–1111, 2003.

[25] K. Umeda, S. Tanabe, and I. Fujita, "Representation of stereoscopic depth based on relative disparity in macaque area V4," *J. Neurophysiol.*, vol. 98, no. 1, pp. 241–252, 2007.

[26] P. Neri, "A stereoscopic look at visual cortex," *J. Neurophysiol.*, vol. 93, no. 4, pp. 1823–1826, 2005.

[27] R. T. Born and D. C. Bradley, "Structure and function of visual area MT," *Annu. Rev. Neurosci.*, vol. 28, pp. 157–189, Mar. 2005.

[28] A. Wong, *Eye Movement Disorders*. Oxford, U.K.: Oxford Univ. Press, 2007.

[29] U. Büttner and J. A. Büttner-Ennever, "Present concepts of oculomotor organization," *Prog. Brain Res.*, vol. 151, pp. 1–42, 2005.

[30] P. D. Gamlin, "Subcortical neural circuits for ocular accommodation and vergence in primates," *Ophthalmic Physiol. Opt.*, vol. 19, no. 2, pp. 81–89, Mar. 1999.

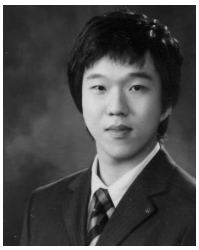
[31] M. S. Banks, A. B. Sekuler, and S. J. Anderson, "Peripheral spatial vision: Limits imposed by optics, photoreceptors, and receptor pooling," *J. Opt. Soc. Amer.*, vol. 8, no. 11, pp. 1775–1787, Nov. 1991.

[32] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated video compression with optimal rate control," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 977–992, Jul. 2001.

[33] W. S. Geisler and J. S. Perry, "Real-time foveated multiresolution system for low-bandwidth video communication," *Proc. SPIE*, vol. 3299, pp. 294–305, Jul. 1988.

[34] A. K. Moorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 193–201, Apr. 2009.

[35] H. Kim, S. Lee, and A. C. Bovik, "Saliency measurement on stereoscopic videos," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1476–1490, Apr. 2014.
[36] (2012). *Standard for the Quality Assessment of Three Dimensional (3D) Displays, 3D Contents and 3D Devices Based on Human Factors*. [Online]. Available: http://grouper.ieee.org/groups/3dhf
[37] L. Goldmann, F. De Simone, and T. Ebrahimi, "A comprehensive database and subjective evaluation methodology for quality of experience in stereoscopic video," *Proc. SPIE*, San Jose, CA, USA, vol. 7526, Jan. 2010, p. 75260S.
[38] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1427–1441, Jun. 2010.
[39] P. T. Fox and M. E. Raichle, "Stimulus rate dependence of regional cerebral blood flow in human striate cortex, demonstrated by positron emission tomography," *J. Neurophysiol.*, vol. 51, no. 5, pp. 1109–1120, May 1984.
[40] Y. Nojiri, H. Yamanoue, A. Hanazato, and F. Okano, "Measurement of parallax distribution and its application to the analysis of visual comfort for stereoscopic HDTV," in *Proc. Stereosc. Displays Virt. Real. Syt. X*, vol. 5006. 2003, pp. 195–205.
[41] J. Choi, D. Kim, S. Choi, and K. Sohn, "Visual fatigue modeling and analysis for steroscopic video," *Opt. Eng.*, vol. 51, no. 1, p. 017206, Feb. 2012.
[42] B. Schölkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, "New support vector algorithms," *Neural Comput.*, vol. 12, no. 5, pp. 1207–1245, 2000.
[43] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discovery*, vol. 2, no. 2, pp. 121–167, 1998.
[44] C. D. Wickens, *Engineering Psychology & Human Performance*, 2nd ed. New York, NY, USA: Harper Collins, 1992.
[45] B. C. Kuo, *Automatic Control Systems*, 7th ed. New York, NY, USA: Wiley, 1995.
[46] I. P. Fawcett, G. R. Barnes, A. Hillebrand, and K. D. Singh, "The temporal frequency tuning of human visual cortex investigated using synthetic aperture magnetometry," *NeuroImage*, vol. 21, no. 4, pp. 1542–1553, 2004.
[47] S. J. Anderson, I. E. Holliday, K. D. Singh, and G. F. A. Harding, "Localization and functional analysis of human cortical area V5 using magneto-encephalography," *Proc. Roy. Soc. London B, Biol. Sci.*, vol. 263, no. 1369, pp. 423–431, Apr. 1996.
[48] W. Lutzenberger, F. Pulvermüller, T. Elbert, and N. Birbaumer, "Visual stimulation alters local 40-Hz responses in humans: An EEG-study," *Neurosci. Lett.*, vol. 183, nos. 1–2, pp. 39–42, Jan. 1995.
[49] H. K. Gomarus, M. Althaus, A. A. Wijers, and R. B. Minderaa, "The effects of memory load and stimulus relevance on the EEG during a visual selective memory search task: An ERP and ERD/ERS study," *Clin. Neurophysiol.*, vol. 117, no. 4, pp. 871–884, Apr. 2006.
[50] H.-C. O. Li, J. Seo, K. Kham, and S. Lee, "Measurement of 3D visual fatigue using event-related potential (ERP): 3D oddball paradigm," in *Proc. 3DTV Conf., True Vis.-Capture, Transmiss. Display 3D Video*, Istanbul, Turkey, May 2008, pp. 213–216.



Taewan Kim received the B.S., M.S., and Ph.D. degrees in electrical and electronic engineering from Yonsei University, Seoul, Korea, in 2008, 2010, and 2015, respectively. He is currently with the Video Technology Laboratory, SK Telecom, Seoul. His research interests include quality assessment of 2D and 3D image and video, 3D video coding, cross-layer optimization, and wireless multimedia communications. He has participated in the IEEE Standard Working Group for 3D Quality Assessment (IEEE P3333.1). He received the Samsung Humantech Thesis Prize in 2013.

Sanghoon Lee (M'05–SM'12) received the B.S. degree in electrical engineering from Yonsei University, in 1989, the M.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology, in 1991, and the Ph.D. degree in electrical engineering from The University of Texas at Austin, in 2000. From 1991 to 1996, he was with Korea Telecom. From 1999 to 2002, he was with Lucent Technologies on 3G wireless and multimedia networks. In 2003, he joined the faculty of the Department of Electrical and Electronics Engineering, Yonsei University, Seoul, Korea, where he is a Full Professor. His research interests include image/video quality assessment, medical image processing, cloud computing, sensors and sensor networks, wireless multimedia communications, and wireless networks. He received the 2015 Yonsei Academic Award from Yonsei University, the 2012 Special Service Award from the IEEE Broadcast Technology Society, and the 2013 Special Service Award from the IEEE Signal Processing Society. He was an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING (2010–2014). He has been an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS (2014-) and the Chair of the IEEE P3333.1 Quality Assessment Working Group (2011-). He currently serves on the IEEE IVMSP Technical Committee (2014-). He was the Technical Program Co-Chair of the International Conference on Information Networking in 2014, and the Global 3D Forum in 2012 and 2013, and the General Chair of the 2013 IEEE IVMSP Workshop. He also served as a Special Issue Guest Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING in 2013, and an Editor of the *Journal of Communications and Networks* (2009–2015).

Alan Conrad Bovik (S'80–M'81–SM'89–F'96) is currently the Curry/Cullen Trust Endowed Chair Professor with The University of Texas at Austin, Austin, TX, USA, where he is also the Director of the Laboratory for Image and Video Engineering, and a Faculty Member with the Department of Electrical and Computer Engineering and the Center for Perceptual Systems, Institute for Neuroscience. He has authored over 650 technical articles in these areas, and holds two U.S. patents. His several books include the recent companion volumes *The Essential Guides to Image and Video Processing* (Academic Press, 2009). His research interests include image and video processing, computational vision, and visual perception.

He was a recipient of a number of major awards from the IEEE Signal Processing Society, including the best paper award (2009), the Education Award (2007), the Technical Achievement Award (2005), and the Meritorious Service Award (1998). He was also a recipient of the Honorary Member Award from the Society for Imaging Science and Technology (2013), the Society of Photo-Optical and Instrumentation Engineers (SPIE) Technology Achievement Award (2012), and the IS&T/SPIE Imaging Scientist of the Year Award (2011). He received the Hocott Award for Distinguished Engineering Research at The University of Texas at Austin. He was a recipient of the Distinguished Alumni Award from the University of Illinois at Urbana–Champaign (2008), the IEEE Third Millennium Medal (2000), and two Journal Paper Awards from the International Pattern Recognition Society (1988 and 1993). He is a fellow of the Optical Society of America, SPIE, and the American Institute of Medical and Biomedical Engineering. He has been involved in numerous professional society activities, including the Board of Governors of the IEEE Signal Processing Society (1996–1998), the Co-Founder and Editor-in-Chief of the IEEE TRANSACTIONS ON IMAGE PROCESSING (1996–2002), an Editorial Board Member of the PROCEEDINGS OF THE IEEE (1998–2004), a Series Editor of the *Image, Video, and Multimedia Processing* (Morgan and Claypool Publisher, 2003–present), and the Founding General Chairman of the First IEEE International Conference on Image Processing in Austin (1994).

Dr. Bovik is a Registered Professional Engineer in the state of Texas, and a Frequent Consultant to legal, industrial, and academic institutions.